K-mean clustering and Curve fitting using logistic growth model

Introduction

The economic output of a nation or region per person is measured by the GDP per capita. The data includes the GDP per capita of some regions of the world over the period of 49 years. The data was gotten from the world bank database.

Aim

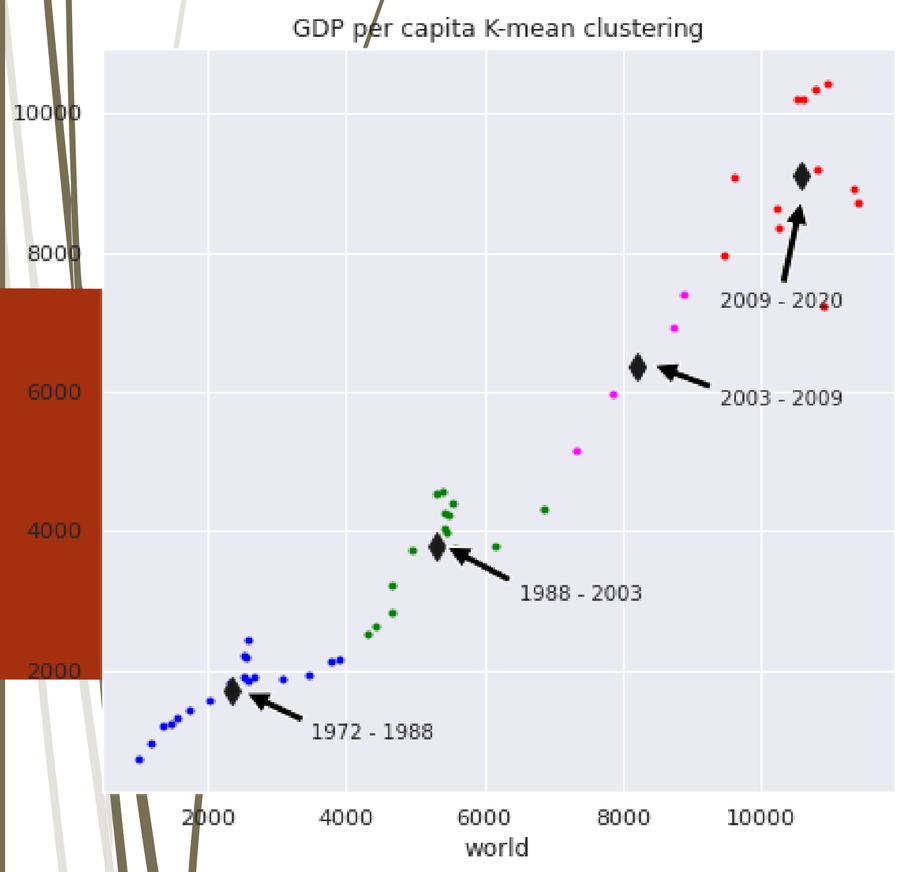
- Using K-mean clustering Algorithm for classification and logistic growth model for predicting the GDP per capita in 10years from 2020

Methods

- Classifying the GDP per Capita of 2 regions over a period of 49 years into 4 Using K-MEAN clustering algorithms from python sklearn library.
- Building a model that fits the data using the scipy.optimize library from python and predicting the GDP per capita of 2 regions in 10 years' time from 2020 with confidence range.

Results

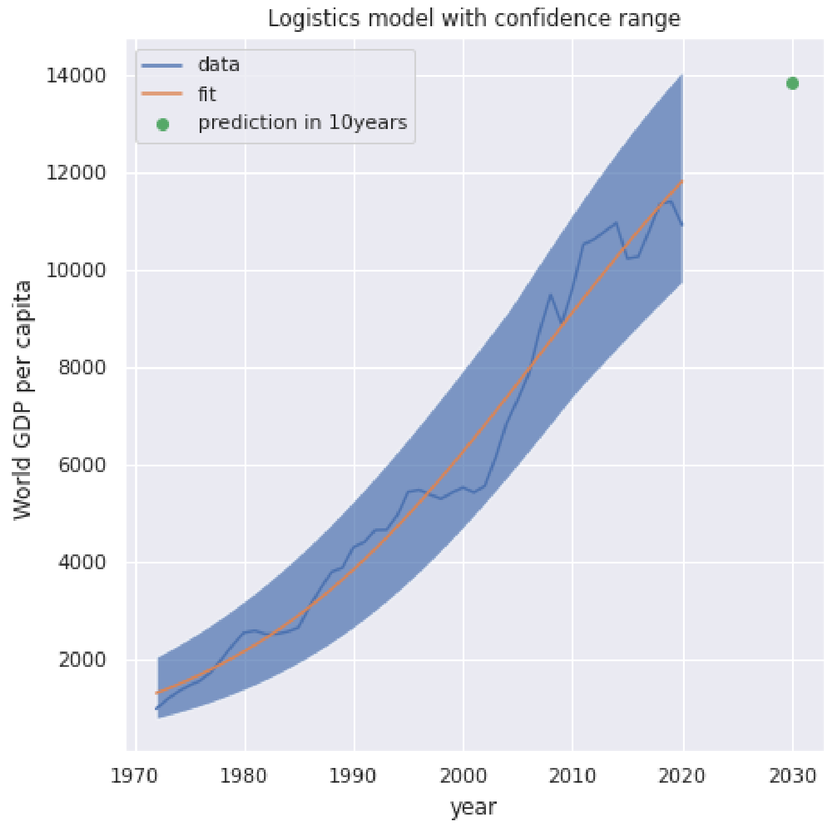
- After using the above method, the following results were obtained



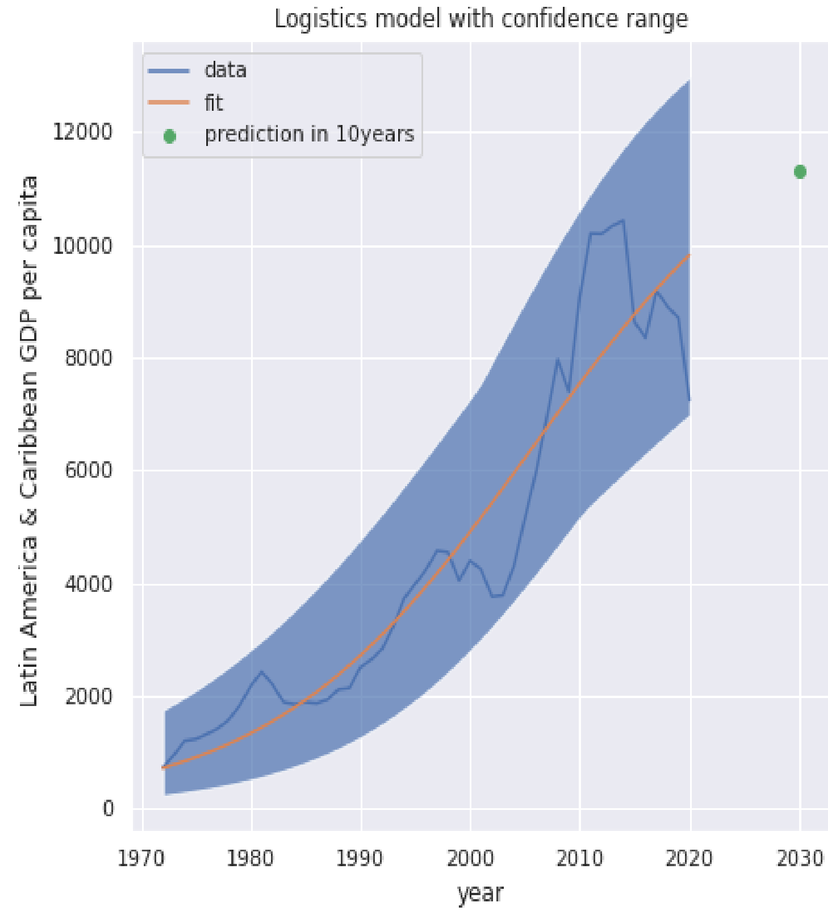
Performing a k mean clustering on the GDP per capital of world vs Latin America and the Caribbean shows how the GDP has grown every decade. It can be observed in this figure that a new decade ushers in a new record high GDP per capita.

- An interesting property of GDP per capita is the trend it follows from year to year. If we can understand the trend it follows, we can build a simple model to predict the GDP per capita of the region in the future.
- Looking at the plot of the GDP per capita of both regions it can be observed that the graph plot looks like a logistics growth. Whereby, the growth rate is rapid at the beginning and then starts to reduce with time.

- The figure below shows how the logistics growth model fits the data. The GDP per capita in 10 years from 2020 can also be predicted by the model.



The figure shows how the logistics growth model we have built fits our world data. The GDP per capita in 10 years from 2020 can also be predicted by the model which is 13837USD. And the confidence range as also been computed as 11680 USD to 16054 USD.



This figure shows how the logistics growth model fits the Latin America & Caribbean GDP per capita data. The GDP per capita in 10 years from 2020 can also be predicted by the model which is 11301USD. And the confidence range as also be computed as 8419 USD to 14159 USD.

- The logistics model fits the World GDP per capita better compared to the Latin America and Caribbean region data. This is because of the non-steady trend in the GDP per capita of the Latin America and Caribbean region. However, the confidence range computed is robust enough to Predict the GDP per capita in 10 years' time from 2020.

References
Data source: https://databank.worldbank.org/
Github link: https://github.com/Ayoola17/k-mean-and-curve-fitting.git